

FORM PTO-1390 (REV 5-93)	U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE	ATTORNEY DOCKET NO. 100021-00073 DATE: February 25, 2002 U.S. APPLN. NO. (IF KNOWN, SEE 37 C.F.R. 1.5) 10/049635
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371		
INTERNATIONAL APPLICATION NO. PCT/JP99/05398	INTERNATIONAL FILING DATE 30 September 1999	PRIORITY DATE CLAIMED
TITLE OF INVENTION: APPARATUS AND METHOD FOR DISCRIMINATING DOCUMENTS		
APPLICANT(S) FOR DO/EO/US: Takayuki MATSUI; Yutaka KATSUMATA; Kazunori YAMAMOTO; Shimichi EGUCHI		
<ol style="list-style-type: none"> 1. <input checked="" type="checkbox"/> This is a FIRST submission of items concerning a filing under 35 U.S.C. 371. (THE BASIC FILING FEE IS ATTACHED) 2. <input type="checkbox"/> This is a SECOND or SUBSEQUENT submission of items concerning a filing under 35 U.S.C. 371. 3. <input checked="" type="checkbox"/> This express request to begin national examination procedures [35 U.S.C. 371(f)] at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1). 4. <input type="checkbox"/> A proper demand for International Preliminary Amendment was made by the 19th month from the earliest claimed priority date. 5. <input checked="" type="checkbox"/> A copy of the International Application as filed [35 U.S.C. 371(c)(2)] <ol style="list-style-type: none"> a. <input checked="" type="checkbox"/> is transmitted herewith (required only if not transmitted by the International Bureau). b. <input type="checkbox"/> has been transmitted by the International Bureau. c. <input type="checkbox"/> is not required, as the application was filed in the United States Receiving Office (RO/US). 6. <input checked="" type="checkbox"/> A translation of the International Application into English [35 U.S.C. 371(c)(2)]. 7. <input type="checkbox"/> Amendments to the claims of the International Application under PCT Article 19 [35 U.S.C. 371(c)(3)] <ol style="list-style-type: none"> a. <input type="checkbox"/> are transmitted herewith (required only if not transmitted by the International Bureau). b. <input type="checkbox"/> have been transmitted by the International Bureau. c. <input type="checkbox"/> have not been made; however, the time limit for making such amendments has NOT expired. d. <input type="checkbox"/> have not been made and will not be made. 8. <input type="checkbox"/> A translation of the amendments to the claims under PCT Article 19 [35 U.S.C. 371(c)(3)]. 9. <input checked="" type="checkbox"/> An oath or declaration of the inventor(s) [35 U.S.C. 371(c)(4)]. 10. <input type="checkbox"/> A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 [35 U.S.C. 371(c)(5)]. <p>Items 11 - 16 below concern other document(s) or information included:</p> <ol style="list-style-type: none"> 11. <input checked="" type="checkbox"/> An Information Disclosure Statement under 37 C.F.R. 1.97 and 1.98. 12. <input checked="" type="checkbox"/> An assignment document for recording. A separate cover sheet in compliance with 37 C.F.R. 3.28 and 3.31 is included. 13. <input type="checkbox"/> A FIRST preliminary amendment. <input type="checkbox"/> A SECOND or SUBSEQUENT preliminary amendment. 14. <input type="checkbox"/> A substitute specification. 15. <input type="checkbox"/> A change of power of attorney and/or address letter. 16. <input checked="" type="checkbox"/> Other items or information: PCT/IB/308; PCT/IB/332; Copy of first page of WO 01/26024; PCT/IB/338; PCT/IPEA/409 and PCT/ISA/210; Certified copy of Japanese Application No. PCT/JP99/05398 dated. 9/30/1999 Drawings (Figs. 1-12; 12 sheets) 		

2

12/pts

- 1 -

SPECIFICATION

APPARATUS AND METHOD FOR DISCRIMINATING DOCUMENTS

CROSS-REFERENCE TO RELATED APPLICATIONS

5 This application is a continuation application and is based upon PCT/JP99/05398, filed on September 30, 1999.

Technical Field

10 The present invention relates to a document discriminating apparatus, and a method for discriminating documents, which are suitable for use in processing documents at financial institutions and, more particularly, to an apparatus and method, for discriminating documents, for use in processing documents
15 such as privately prepared slips in various formats.

Background Art

In recent years, image data readers such as optical character readers (OCRs) have been developed as devices for reading character information as image data for
20 character recognition, and image data readers are now widely used in various industries to attempt to perform jobs efficiently.

For example, operators at windows of financial institutions attempt to perform their jobs more
25 efficiently by processing documents more efficiently using the image data readers.

In particular, in order to increase efficiency in performing jobs such as processing the documents, it is required not only to process a large number of documents
30 of the same type but also to automatically process a large amount of documents of various formats.

To cope with this, document processing devices provided with an image data reader are used. The image data reader of the document processing device for reading
35 image data on a document is adapted to read image data based on a control from an electronic computer. In addition, for example, image scanners and facsimile

machines are used as the image data reader. Furthermore, the image data reader can be an image data read and recognition device which can both read image data and recognize characters.

5 In addition, the electronic computer functioning as a controller for controlling the image data reader is constituted by an input means which is a keyboard or a mouse for inputting instructions from the operator and data, a computer main body and a display for displaying
10 data or control information. In addition, recognition of image data read by the image data reader is performed by the electronic computer main body.

 Furthermore, the document processing device is provided with a hard disk, which is connected to the
15 electronic computer and stores in advance position information of character data to be recognized for each type of document and information designating types and numbers of characters (hereinafter, referred to as "definition information").

20 Next, an operation will be described which is to be performed when the document processing device is used.

 In recognizing character data described on, for example, an "ELECTRIC RATE BILL" using the image data reader, firstly, the operator operates the keyboard and
25 designates definition information B corresponding to the type of a document (in this case, the document is regarded as an electric rate bill (Document B)).

 Following this, at the electronic computer, the hard disk is accessed to draw the designated definition
30 information B for the document, and the image data reader is notified of the information.

 As this occurs, at the image data reader, reading of image data and recognition of characters can be performed, based on the definition information B which is
35 a control information notified from the computer.

 However, in a method for processing documents, as definition information is designated for each document to

be read through the designation by the operator, the operator has to bear an additional work load, and there may occur designation errors with an increase in the number of pieces of definition information. Moreover, in
5 a case where the operator is required to process thousands of types of documents, the designation by the operator is practically impossible.

To cope with this, there has been proposed a method for automatically reading documents without any
10 designation by the operator, as described above, by describing in advance an specific ID number at a predetermined position on each type of document for discriminating the document of the type from others.

According to this method, in reading the image data
15 of the document by the image data reader, the character recognition is made possible by firstly identifying the ID number put at the predetermined position and then using definition information (in this case, B) corresponding to the ID number.

However, in case the position on an optical reader
20 portion of the image data reader where a document is set is changed in reading image data, for example, even in case document identical to the document whose definition information has already been stored in the hard disk,
25 since a coordinate from the reference point (a physical origin) for an image such as a character data area and a graphic area does not match the coordinate in the definition information, it is determined that they are not in the same layout.

At the document processing device, there may be a
30 case in which the character recognition process is not performed properly in case the character recognition of image data is implemented after the layout of the read image data has been made to match the layout of the
35 definition information. Due to this, firstly, the reference points for the respective images are extracted, the coordinates of the respective images from the

reference points are compared to each other, and whether or not the layouts match each other is determined.

Here, there are the following two methods for extracting a reference point for image data. As
5 preconditions, a document to be read is a pre-printed document, and the printed position of the document on the form is controlled with high accuracy.

In a first method, in a case where image data of a document is read with an image data reader which can
10 discriminate an end face of the document to be read from the background of the form to be read, for example, an upper left-hand side corner portion at the end face of the form is regarded as the reference point.

In addition, in a second method, in a case where
15 image data of a document is read with an image scanner or a facsimile machine, since the read background cannot be discriminated from the end face of the form, a reference mark is printed in advance on a document form to be read, and this reference mark is extracted from the image data
20 to become the reference point.

In this second method, since the reference point position is printed in advance as the reference mark, there is provided an advantage that even in case the position where a document is set on the reader is
25 changed, the reference point can be extracted in a stable fashion.

However, in the document processing device, even in a case where the reference point is extracted accurately using the aforesaid method, since the character
30 describing direction of image data does not become a correct direction in case the reading direction of image data is not the right direction, no character recognition process can be implemented.

For example, in case a money transfer request slip
35 in which information is described horizontally is read from a wrong direction by the image data read and recognition device, the image data of the money transfer

request slip so read is displayed on the display in the wrong direction.

Due to this, in recognizing characters by the document processing device, the operator determines whether or not the reading of the document is implemented from the right direction by looking at the display, and in case the reading direction of the document is determined to be wrong, the operator inputs from the keyboard a designation to rotate the image data of the document so read through 90 to 180 degrees, and a rotational correction process for the image data needs to be implemented so that the image data of the document can be displayed on the display in the right direction.

Incidentally, in executing jobs of transferring monies to accounts at banks, conventionally the operator inputs through the keyboard information on account number, names and amount of money which are described on a document which is a money transfer request slip, and as this occurs, the operator performs the input operation while watching the document and the display in an alternate fashion.

However, when the operator performs the input operation while watching the document and the display alternately, since the movement of the eyes occurs frequently, there exists a problem that the visibility is deteriorated and hence there may be caused an error in matching an item on the document with an item on the display.

To cope with this, there has been proposed a document processing device for reading a document with an image data read and recognition device and displaying the results of character recognition of the image data so read, as well as the image data itself on the display.

According to this document processing device, since information described on the document can be seen directly on the display, the movement of the eyes is reduced, whereby the occurrence of errors in matching

between the character recognition results and the image data can be reduced when confirming and modifying the character recognition results.

5 On the other hand, in recent years, another method has become the main stream, in data processing jobs at financial institutions, in which a client-server system is used to do data processing collectively, at high speeds, at the server.

10 For example, it is proposed that clients, set at respective branches (sales points), and a server, set at a regional center of a financial institution, are connected via exclusive lines or public lines to form a client server system, so that document processing is implemented collectively by the server to thereby attempt
15 to increase the efficiency of the document processing jobs.

In batch processing the documents by the server as described above, since the amount of data collected at the server becomes gigantic, it becomes impossible for
20 the operator to do document processing at the server, and therefore, a system must be designed in which the server can automatically do document processing without the involvement of the operator.

To this end, the application of a document
25 discriminating process using documents on which ID numbers are described to this client server system makes it possible for the server to automatically discriminate the types of documents and to read documents.

In addition, also at clients of this client server
30 system, as has been described above, the results of character recognition of the image data read by the image data reader, as well as the image data itself can be displayed on the display, whereby the information described on the documents can be seen directly on the
35 display, thus making it possible to reduce the occurrence of errors in matching the character recognition results and the image data when confirming and modifying the

character recognition results.

5 However, in the aforesaid method for processing documents, only documents prepared exclusively for the document processing device can be read by the device, and general documents which have conventionally been in use cannot be used. Therefore, there exists a problem that special documents have to be prepared for use with such a document processing device.

10 Additionally, with the first reference point extracting method described above, while the print position of the document on the form needs to be controlled with high accuracy, in the case of a document printed with, for example, a word processor, papers are set manually, and the print position tends to vary every
15 time printing is carried out, this causing the problem that using the upper left-hand side corner at the end face of the form as the reference point is not suitable for use with the word processor.

20 Furthermore, with the second reference point extracting method described above, the document to be read is the special document on which the reference mark is printed, this causing the problem that the reference point cannot be extracted with respect to general documents on which no reference mark is printed.

25 In addition, even with a method in which a special point on the front side of the document is used as the reference point, this method being adopted in the recognition technology used on document readers, there is also caused a problem that the method cannot function
30 effectively if the layout of a document to be read cannot be identified to some extent.

35 Additionally, even in case a similar document as that which has been read is read again, there occurs a case in which an image which has been read does not match the image which was read before due to dust or a thin spot, and when this happens there may be a case where the same reference point cannot be extracted.

Furthermore, in order for the rotational correction processing to be automatically implemented at the server, the character recognition of the image data which is actually read is implemented, and whether the reading direction of the document is correct or not is determined by whether or not the character recognition is possible, this causing a problem that the increase in efficiency of the document processing job is remarkably reduced.

In addition, in the aforesaid method in which the character recognition results are confirmed and modified by displaying on the display the character recognition results, as well as the image data itself, since matching between the image data and the confirmation items is visually implemented, in case there are many items to be confirmed in the image data, there is also provided a problem that it is not possible to eliminate recognition errors in case where there are many items in the image data to be confirmed.

Furthermore, in a case where image data cannot be displayed on the display at one time, in order to refer to data disposed before, the display screen has to be scrolled down and, as this occurs by depressing a predetermined key with the finger the scroll operation, this results in a more complicated operation.

Disclosure of the Invention

The present invention was made in view of these problems, and an object thereof is to provide a document discriminating apparatus for processing documents which are general documents that have been conventionally used and which have various types of formats like privately prepared documents, and a document discriminating method.

In addition, another object of the present invention is to implement at all times stably and automatically the extraction of a reference point on image data of a printed document even for a document printed on a normal type of paper with a word processor without using the conventional end face of the form and a reference mark.

A document discriminating apparatus according to the present invention, which can attain the objects, comprises an image reading means for reading image data from a document prepared in an optional format, an image data cutting out means for cutting out data corresponding to a designated special portion of the document from the image data read by the image reading means, a color constituent extracting means for analyzing the color constituents of the image data cut out by the cutting out means and setting a color separation parameter for the specific color constituent, and a color constituent separating means for producing data information for the specific portion from the image data cut out based on the color separation parameter from the color constituent extracting means.

Here, the color constituents are analyzed by the three primary colors of the color, and one of the three primary colors of the color is selected as the specific color constituent. The color separation parameter related to the color constituent of the specific color is determined based on the concentration distributions of the three primary colors of the color.

Then, the document is discriminated by cutting out image data corresponding to a plurality of designated specific locations from the read image data, preparing data information for the specific portion from the cut out image data based on the color separation parameters, and comparing the data information with the data information stored in the document discriminating dictionary unit.

Consequently, according to the document discriminating apparatus according to the present invention, since accurate data information can be prepared for the cut out image data when automatically discriminating the type of the document from which image data is read with the image data read and recognition device, even if there exist a plurality of documents to

be read with the image reader in a mixed fashion, the operator can process the document without paying attention to the definition of each document to thereby attempt to increase the efficiency of the job. In addition, there is no need to describe an ID number for discriminating the document itself, whereby general documents can be used, thus making it possible to deal with the existing systems with ease.

Brief Description of the Drawings

Fig. 1 is a typical diagram showing the construction of a document discriminating apparatus,

Fig. 2 is a block diagram showing a document discriminating apparatus which constitutes the base of the present invention,

Fig. 3 is a diagram for explaining the operation of the document discriminating apparatus which constitutes the base of the present invention when a registration step is activated,

Fig. 4 is a control block diagram of the document discriminating apparatus which constitutes the base of the present invention when attention is paid to the activation of the registration step,

Fig. 5 is a flowchart for explaining the operation of the document discriminating apparatus which constitutes the base of the present invention when the registration step is performed,

Fig. 6 is a control block diagram of the document discriminating apparatus according to the present invention when attention is paid to the activation of a document determination step,

Fig. 7 is a flowchart for explaining the operation of the document discriminating apparatus which constitutes the base of the present invention when attention is paid to the activation of the document determination step,

Fig. 8 is a diagram showing an example in which color constituent extraction and separation parameters

are determined,

Fig. 9 is a control block diagram of an embodiment of the present invention when attention is paid to the activation of the registration step of the document discriminating apparatus,

Fig. 10 is a flowchart for explaining the operation of the embodiment of the present invention when attention is paid to the activation of the registration step of the document discriminating apparatus,

Fig. 11 is a control diagram of another embodiment of the present invention when attention is paid to the activation of a document discriminating step of a document discriminating apparatus, and

Fig. 12 is a flowchart for explaining the operation of the embodiment of the present invention when attention is paid to the activation of the document discriminating step of the document discriminating apparatus.

Mode for Carrying out the Invention

To clarify the effectiveness provided by the present invention, firstly, the construction of a document discriminating apparatus which constitutes the base of the present invention will be described.

Fig. 1 is a functional block diagram showing the overall construction of a document discriminating apparatus. The document discriminating apparatus comprises an image data reader 101 for reading image data on a document, an electronic computer 101 for controlling the reading operation of image data, a hard disk 103 connected to the electronic computer 102 for storing in advance position information of character data to be recognized for each type of document and information designating the type and number of characters (hereinafter, referred to as "definition information"), a display 104 for displaying data and control information and an input means 105 such as a keyboard or a mouse for inputting instructions from an operator and data. Then, the image data reader 101 is adapted to read image data

from a document 106 such as an "ELECTRIC RATE BILL."
Note that the image data reader 101 includes for example
an image scanner and a facsimile machine as an image data
reading portion.

5 In the document discriminating apparatus constructed
as described above, in order to deal with the aforesaid
problems, there is proposed a document discriminating
apparatus for processing documents which are general
documents that have been conventionally used and which
10 have various types of formats like privately prepared
documents.

 In addition, there is proposed a document
discriminating apparatus adapted to implement at all
times stably and automatically the extraction of a
15 reference point on image data of a printed document even
for a document printed on a normal type of paper with a
word processor without using the conventional end face of
the form and reference mark.

 Then, referring to the appended drawings, the
20 proposed document discriminating apparatus will be
described below.

 Fig. 2 is a block diagram showing the document
discriminating apparatus, and similarly, the document
discriminating apparatus shown in Fig. 2 also comprises
25 an image data reader 101, an electronic computer 102
(this electronic computer 102 comprising, as will be
described, an input unit 105, a display 104 and a control
unit 201) and a hard disk 103.

 Here, the image data reader 101 is designed to read
30 image data of a document and, as with the image data
reader shown in Fig. 1, an optical character reader (an
OCR) and an image scanner can be used as the image data
reader.

 In addition, the control unit 201 is designed to
35 process the image data of a document read with the image
data reader 101 as the document data based on the image
data so read, and can be constituted by a CPU and a

memory of the electronic computer 102.

5 Additionally, connected to this control unit 201 are the input unit 105 such as a keyboard or a mouse for inputting data or instructions from an operator thereinto and the display 104 for displaying thereon the image data read with the image data reader 101.

 Furthermore, the hard disk (a file memory) 103 is designed to store the image data of all documents that are read with the image data reader 101.

10 Incidentally, as shown in Fig. 2, the control unit 201 comprises as functional blocks an image data storing memory 202, an image data cutting out unit 203, a document discriminating dictionary unit 204, a data comparison unit 205, a threshold setting unit 206, a
15 document determination unit 207, a definition storing unit 208, a definition storing table 211, a character recognition unit 209 and a character recognition result storing unit 210.

 The image data storing memory 202 is designed to
20 temporarily store the image data of a document read with the image data reader 101. The image data cutting out unit 203 is designed to have a function as a document discriminating information extraction means for
25 extracting required document discriminating information described on a document from the image data of the document stored in the image data storing memory 202 when receiving a designation of information (discriminating information) to be extracted through the operation of the input unit 105 by the operator.

30 Here, in the image data cutting out unit 203, in extracting required document discriminating information from the discriminating information, the image data of the document read with the image data reader 101 is designed to be displayed on the display 104, and the
35 operator can designate discriminating information based on image graphics displayed on the display 104.

 Note that the operator can designate every described

information such as character information, marks, seals or rules described on a document as information (unique information) to be extracted at the image data cutting out unit 203. Then, coordinate position information for the designated information, size information of the described information and data information are designed to be automatically extracted as document discriminating information through, for example, software processing or firmware processing at the image data cutting out unit 203.

Furthermore, the document discriminating dictionary unit 204 is designed to register document discriminating information extracted at the image data cutting out unit 203 as document discriminating information for a specific document.

To be specific, as shown in Fig. 3, document discriminating information for a document type A having an ID number '0101' affixed thereto is designed to be stored in an area 204a, whereas document discriminating information for a document type B having an ID number '0102' affixed thereto is designed to be stored in an area 204b. Thus, like information is stored sequentially in accordance with ID numbers.

Here, an operation will be described below in which document discriminating information is registered as document discriminating information for a specific document. In Fig. 3, a state is shown in which the registration of the document type A has been completed, and a document type B of an "ELECTRIC RATE BILL" is about to be registered next. A document 106 which is the document type B read with the image data reader 101 is displayed on the display 104. The operator designates a plurality of locations which can be features to discriminate the document 106 (unique information) via the input means 105.

In the figure, a state is shown in which the 'electric rate' described on the document 106 is

designated as unique information D1 and 'FUJI ICHIRO' as
unique information D2. Position information (X_0 , Y_0),
size information and data information such as characters
are read with respect to the 'ELECTRIC RATE' and the
5 pieces of information are stored in a document
discriminating information 1 column of the area 204b
which is a storing place for the document type B in the
document discriminating dictionary unit 204 as a single
piece of unique information. Then, as with the unique
10 information D2, 'FUJI ICHIRO' the information is stored
in a document discriminating information 2 column of the
area 204b. Thus, the plurality of pieces of unique
information on the feature portions required to specify a
single document type are stored.

15 Consequently, an operation as a registration step is
designed to be carried out by the image data storing
memory 202, the image data cutting out unit 203 and the
document discriminating dictionary unit 204 in which the
document discriminating information described on the
20 specific document is extracted from the image data of the
specific document read with the image data reader 101 and
the document discriminating information is registered in
the document discriminating dictionary unit 204.

25 Note while the image data of the document read with
the image data reader 101 is temporarily stored in the
image data storing memory 202 when the document
discriminating information is registered in the document
discriminating dictionary unit 204, image data on the
whole document is designed to be stored in the hard disk
30 103.

In addition, the data comparison unit 205 has a
function as a verification means for verifying whether or
not document discriminating information registered in the
document discriminating dictionary unit 204 exists in the
35 image data of a specific document read out from the image
data storing memory 202. Furthermore, the data
comparison unit 205 has also a function as a reference

means for referring whether or not document
discriminating information registered in the document
discriminating dictionary unit 204 exists in the image
data of an optional document read out with the image data
5 reader 101 and stored in the image data storing memory
202.

Furthermore, the document determination unit 207 has
a function as a determination means for determining
whether or not a specific document has entirely been
10 specified as a whole by determining whether or not the
specific document can be recognized based on the results
of the verification at the data comparison unit 205.
Then, the document determination unit 207 also has a
function as a document discriminating means for
15 discriminating whether or not an optional document is the
specific document based on the result of referring at the
data comparison unit 205 functioning as the reference
means.

Here, to be specific, the data comparison unit 205
20 calculates the degree of matching for image data inputted
from the image data reader 101 by collating the
information extracted at the image data cutting out unit
203 and corresponding document discriminating information
from the document discriminating dictionary unit 204.
25 Then, the document determination unit 207 functioning as
the document discriminating means is designed to
determine whether or not a document for the image data
inputted from the image data reader 101 can be
discriminated by comparing the degree of matching of the
30 document discriminating information from the data
comparison unit 205 with a threshold from the threshold
setting unit 206.

Consequently, an operation as a verification step
for determining whether or not the discrimination of a
35 specific document can be implemented is performed by the
document discriminating dictionary unit 204, the data
comparison unit 205, the threshold setting unit 206 and

the document determination unit 207.

5 In addition, in determination by the document determination unit 207 of the degree of matching based on the threshold information from the threshold setting unit 206, the threshold information from the threshold setting unit 206 is designed to be set such that a determination can be made by absorbing an error that would occur during the reading operation by the image data reader 101 and a printing error found on the document itself.

10 In addition, the definition storing unit 208 is designed to read out definition information for recognizing data described on the document from the definition storing table 209 to temporarily store the information therein. The storing of the information is
15 implemented when a document read with the image data reader 101 can be recognized as the specific document that has already been registered at the document discriminating dictionary unit 204 by the document determination unit 207 when the system is operated.

20 The definition storing table 209 is designed to hold the definition information (for example, read position information, character attribute information, the number of characters or numbers to be read) for character recognizing the contents described on a specific document
25 corresponding to the document discriminating information registered at the document discriminating dictionary unit 204.

In addition, the character recognition unit 210 is designed to character recognize the image data in
30 accordance with the definition information when the definition information corresponding to the image data from the definition storing unit 208 is inputted thereinto with respect to the image data stored in the image data storing memory 202 as being related to the
35 document which has been able to be recognized as the specific document that has been registered at the document discriminating dictionary unit 204.

Furthermore, the character recognition result storing unit 211 is designed to store character information recognized at the character recognition unit 210.

5 The operation of the document discriminating apparatus constructed as has been described heretofore will be described separately as in a registration step and in a document discriminating step.

[Processing in Registration Step]

10 First, the operation of the document discriminating apparatus in the registration step will be described below with reference to a control block diagram shown in Fig. 4 in which attention is paid to the activation of the registration step, a flowchart shown in Fig. 5
15 explaining the operation performed at the time of activation of the registration step and Fig. 3.

Namely, as shown in Fig. 3, when image data of a document 106 (for example, an electric rate bill) is read at the image data reader 101 through the operation of the
20 operator (step S501), the image data so read is temporarily stored at the image data storing memory 202 (step S502), and this image data is also stored in the hard disk 103, whereby the whole image data read at the image data reader 101 is stored (step S503).

25 Note that as shown in Fig. 3, the image data read at the image data reader 101 is displayed via the display 104 (step S504).

Here, in case the image data stored in the image data storing memory 202 and the hard disk 103 is first
30 image data on a document that has ever been read, the document discriminating information is stored in the document registration dictionary unit 204 as will be described below.

Namely, a plurality of pieces of unit information to
35 be extracted are designated to the image data cutting out unit 203 when the operator operates the input unit 105 while referring to the display 104 (step S505).

At the image data cutting out unit 203, when the unique information described on the document is designated, position information, size information and data information related to the unique information are extracted automatically from the image data of the document stored in the image data storing memory 202 (step S506) and are then registered in the document discriminating dictionary unit 204 as the document discriminating information (step S507).

While the case shown in Fig. 3 has been described before, for example, by operating the input unit 105 the operator designates an electric rate showing the details of an amount of money to be paid which is described on the bill as first unique information D1 and 'FUJI ICHIRO' indicating the designation of a payer as second unique information D2. Then, the position information, size information and data information in the first information are stored whereas the position information, size information and data information in the second information are extracted at the image data cutting out unit 203. Furthermore, in case the operator designates a plurality of unique information, the information corresponding to the unique information is sequentially extracted at the image data cutting out unit.

As a result, the extracted document discriminating information for the document type B is stored in the area 204b of the document discriminating dictionary unit 204 as the document discriminating information of the document type B having the ID number '0102' affixed thereto.

Note that in this document discriminating apparatus, the image data cut out at the image data cutting out unit 203 is used only for discriminating the document.

In addition, in this document discriminating apparatus, it is ensured that the document can be discriminated in the verification step and the document discrimination step by registering the plurality of

pieces of document discriminating information per single document without performing a normalization process of image data.

[Operation in Document Discrimination Step]

5 Next, the operation of the document discriminating apparatus according to the embodiment of the present invention in the document discriminating step will be described below with reference to a control block diagram shown in Fig. 6 in which attention is paid to the
10 activation of the document discriminating step, and a flowchart shown in Fig. 7 explaining the operation performed at the time of activation of the document discriminating step.

15 In the verification step, whether or not the document can be specified using the document discriminating information registered at the document discriminating dictionary unit 204 can be verified with respect to the images of all the documents stored in the hard disk 103. When the verification is completed, when
20 an actual document discrimination is implemented, an operation described below as a discrimination step for specifying the type of the document will be performed for the image data of an optional document.

25 Namely, when the image data of a certain document is read through the operation by the operator of the image data reader 101 (step S701), the image data so read is then temporarily stored in the image data storing memory 202 (step S702).

30 Next, at the image data cutting out unit 203, a plurality of pieces of image data (discriminating information) are extracted and cut out based on the position information and size information which constitute the document discriminating information in the document type sequentially selected from the document
35 discriminating dictionary unit 204 for the image data temporarily stored in the image data storing memory 202 (steps S703, S704).

Following this, at the data comparison unit 205, a comparison determination is implemented by calculating the degree of matching between the data information of all the image data cut out at the image data cutting out unit 203 and the data information constituting the document discriminating information (step S705).

Furthermore, at the document determination unit 207, whether or not the image data read with the image data reader 101 can specify the type of the document with the document discriminating information from the document discriminating dictionary unit 204 by comparing the degree of matching calculated as the comparison determination result from the data comparison unit 205 with the determination standard of the degree of matching set at the threshold setting unit 206 (step S706).

In addition, in making a determination at the document determination unit 207, in case any of the plurality of pieces of image data cut out at the image data cutting out unit 203 does not match the document discriminating information from the document discriminating dictionary unit 204, it is determined as a different type of document.

To be specific, in step S706, for example, in a case where the image data of the first document type A is read from the image data reader 101, in case the document discriminating information sequentially read out from the document discriminating dictionary unit 204 to be compared with is related to the first document type A, it is determined as matched with respect to the document discriminating information (Y), and the read image data is specified as the matching document type. As this occurs, the specified document type is recorded in a memory, not shown, within the control unit 201 (step S707).

In addition, while the image data of the first document type A is read in from the image data reader 101, for example, in case the document discriminating

information sequentially read out from the document discriminating dictionary unit 204 as an object for comparison is related to the second document type B, it is determined as unmatched (N). As this occurs, as the document type cannot be specified, there is no need to record the result, then the flow proceeds directly to step S708 by bypassing step S707. Then, in case the degree of matching has not been determined with respect to all the types of documents in the document discriminating dictionary unit 204 (N), then the flow returns to step S703, where the degree of matching is determined again using the document discriminating information related to other types of documents from the document discriminating dictionary unit 204.

Thereafter, as in the case with the aforesaid examples, the degree of matching of the document discriminating information is determined with respect to the image data read with the image data reader 101 based on the document discriminating information for each type of document stored in the document discriminating dictionary unit 204 (step S708).

Here, in case the image data read with the image data reader 101 can be specified as the single type of document in step S709 (Y), the specified type of document is then notified to the operator, and the type so specified is outputted to the definition storing unit 208 (step S710).

On the contrary, in case the image data cannot be specified as the single type of document (N), the operator is notified that the document cannot be specified by displaying, for example, a message to that effect on the display 104 (step S711).

In addition, at the definition storing unit 208, when the document type which is specified as the single type is inputted, the definition information (read position information, character attribute information, the number of characters and/or numbers to be read)

corresponding to the specified type of document is read out from the definition storing table 211.

5 As a result, at the character recognition unit 209, this definition information and the character information described on the document whose image data is read with the image data reader 101 which is stored in the image data storing memory 202 are recognized and the character information as the result of recognition is stored in the character recognition result storing unit 210.

10 Consequently, in performing the discriminating step, the image data of an optional document is read with the image data reader 101, and whether or not the optional document is the specific document is discriminated by referring whether or not the document discriminating
15 information registered in the document discriminating dictionary unit 204 exists in the image data of the optional document, whereby character recognition can be implemented.

20 Thus, the aforesaid document discriminating apparatus is constructed as comprising the image data reader 101, the image data storing memory 202, the hard disk 103, the image data cutting out unit 203, the document discriminating dictionary unit 204, the data comparison unit 205 and the document determination unit
25 207. According to the construction, in discriminating documents, the type of document whose image data is read with the image data reader 101 can automatically be discriminated. Then, even if there are a plurality of types of documents to be read with the image data reader
30 101 in a mixed fashion, the operator can process the documents without paying attention to the definition of each document, thereby making it possible to increase the efficiency of the document processing job. Furthermore, since no ID number for discriminating a document itself
35 needs to be described, general types of documents can be used, and hence the document discriminating apparatus of the invention can be applied to an existing system

without any difficulty.

In addition, when registering the document discriminating information to the document discriminating dictionary unit 204, since required document discriminating information can automatically be taken in through designation by the operator who can do it while looking at the image data of the documents to be registered which are displayed on the display 104, the preparation of the document discriminating dictionary can be facilitated, thus making it possible to attempt to increase the efficiency of the document processing job.

Additionally, the document discriminating information extracted when the operator designates the discriminating information can be designated at the plurality of locations and the document can be specified more accurately by designating the plurality of locations of the document than by designating a single location.

Thus, according to the proposed document discriminating apparatus, the discrimination can also be performed on documents of various types of formats such as privately prepared slips.

However, many of the private documents that have been in use in recent years are colored. A logo mark, which is one of features suitable for discriminating documents, is colored in many cases. In addition, documents of the same format are sometimes differently colored for different applications.

Of course, with the aforesaid document discriminating apparatus, document discriminating information can be obtained which designates the colored portion as a feature for discriminating the document. However, with the above document discriminating apparatus, in registering the plurality of pieces of document discriminating information on each document, even if the colored portion is designated, the discriminating information obtained from the designated colored portion is processed as a monochromatic

information because the computer processes information in a binary fashion, and the monochromatic features of the document are then compared with each other. Due to this, the color information is substituted by monochromatic information, and therefore, the effectiveness of the colored features is deteriorated.

In addition, in a case where the background of a document is colored, the contrast is lowered and the extraction accuracy of the discriminating information is also lowered. Moreover, in the case of a color printed document, tones tend to be changed easily due to printing and/or reading errors, and blending also occurs. In addition, the colored document is also subjected to thin spots and blurs. Due to this, comparison of monochromatic images lowers the discriminating accuracy, and furthermore, even if digitized color images are simply compared with each other, the difference becomes larger and no accurate discrimination can be implemented.

Due to this, even if the color information has a feature in discriminating documents, the colored portion cannot be used as effective discriminating information. If the color information can be used in discriminating documents, it is clear that the discriminating capability can be improved extremely.

To cope with this, according to an embodiment of the present invention, there is provided a document discriminating apparatus which can discriminate documents stably and accurately even if there occurs a change in color through a construction in which data information for data discriminating information for use in discriminating documents can be extracted based on color information.

Next, referring to Fig. 8, a principle of extracting discriminating information based on color information according to the embodiment of the present invention will be described below.

In general, it is well known that a color comprises

the three primary colors when it is separated into constituents. Also in the present invention, the principle of the three primary colors is adopted. However, for the sake of a simple description, in Fig. 8, only two color constituents are shown as a matter of conveniences.

In Fig. 8, the axis of abscissa denotes, for example, a red constituent R and indicates that the density of the color increases as it goes farther in a direction indicated by an arrow R, whereas the axis of ordinate denotes a green constituent G and indicates that the density of the color increases as it goes upper in a direction indicated by an arrow G. In Fig. 8, while no blue constituent is shown, the blue constituent can be indicated in a direction normal to the surface of the paper on which Fig. 8 is illustrated. Therefore, the center 0 denotes white and indicates that as the density of the respective constituents increases, the color becomes closer to black. However, in the figure, the respective axes are shown on different scales.

As has been described with respect to Fig. 3, in registering a document, the document 106 is read with the image data reader 101, the image of the document is displayed on the display 104, feature portions of the document are designated with the input unit 105, and the document discriminating information is extracted. As this occurs, the results of reading with the image data reader 101 may be fully colored on the screen or only the result of reading of the extracted portion may be colored.

In Fig. 8, a state is shown in which, for example, the area of the 'ELECTRIC RATE' on the document type B is designated as an object for extraction, and the image data of the designated extracted portion is read. When analyzing the respective color constituents from the image data, they are represented as a plurality of dots in the figure. Since the object for extraction is a

document, when analyzing the color constituents thereof, the result is a distribution of discontinuity. The figure illustrates a state in which the object for extraction is multi-colored.

5 In the density distribution for each constituent shown in Fig. 8, when looking at the green constituent G and the red constituent R, a group g of dots indicated by a broken line has a green constituent G which is remarkably denser than those in the other groups r_1 , r_2 and r_3 , whereby the group g can be discriminated from the other groups. Namely, the color constituents of the group g are closer to the color image of the extracted portion.

10 Then, in the present invention, the fact that the group g can be discriminated from the other groups r_1 , r_2 and r_3 is used. For example, intermediate points between the group g and the other groups r_1 , r_2 and r_3 are obtained, respectively, and a solid line a passing through these intermediate points is drawn. Here, assuming that this solid line a is a boundary, the whole area is divided into a green constituent G side which is an area A and a non-green constituent side which is an area B. This boundary is made to be a color separation parameter. When the green constituent G is analyzed from the extracted image data by setting the color separation parameter for the green constituent G, in case the green constituent exists in the area A, it is determined that the data exist, and it is adopted for discriminating information as data information equivalent to the monochromatic data. On the contrary to this, in case the green constituent exists in the area B, it is not adopted as discriminating information. Therefore, in order to use the color information for discrimination, the separated color constituents are designated, and the designated color constituents are then stored in the discriminating information dictionary area 204b as conditions for specifying the area A, or the boundary a

is stored as the separation parameter in the discriminating information dictionary area 204b.

To this end, in comparing color images with each other, they cannot be discriminated from each other in case the colors coincide with each other completely. However, even if there occurs a printing or reading error, by designating the specific color constituent and setting the boundary a, in case the specific color constituent exists in the area A separated by the boundary a, since it is regarded as the discriminating information which is made to be monochromatic data, the feature of the extracted portion can sufficiently be specified, and the influence of variation in tone, thin spots or blur can be eliminated. Moreover, there exists no relationship with the background color on the data.

While the discrimination based on the color information has been described, the aforesaid extraction principle can be applied even to a monochromatic object for extraction. Namely, according to the principle of the three primary colors of color, white represents a state in which the three primary constituents of color do not exist at all, and on the contrary, black represents a state in which the three primary constituents exist to their maximum level. Then, in a case where the extracted portion is monochromatic, since it is clear that the results of color constituent extraction show that the density distributions of the three primary constituents are in the vicinity of their maximum values, respectively, in case any of the three primary constituents is selected and designated, data can be extracted as discriminating information.

Thus, even if the object for extraction contains color and monochromatic information in a mixed fashion, as is described above, it is ensured that discriminating information including monochromatic information can be obtained by setting the color separation parameter related to the designated color constituent.

Thus, while the extraction principle of color information has been described heretofore, the designation of color constituent and setting of the separation parameter are implemented while the results of the extraction of the extracted portion are indicated on the diagram shown in Fig. 8 which is indicated on the screen of the display 104. Then, extraction of color information does not have to be implemented on a new document, and as long as the information is good enough to be read with the image data reader 101, color information can be extracted on a used document and can be registered in the discriminating information.

Next, referring to Figs. 9 to 12, a document discriminating apparatus will be described below which adopts the aforesaid color extraction principle according to the embodiment of the present invention.

While the basic construction of a control unit of the document discriminating apparatus according to the embodiment of the present invention is similar to the block construction shown in Fig. 2, the basic construction according to the embodiment differs from that shown in Fig. 2 in that the control unit further has a color constituent extracting unit and a color constituent separating unit for specifying the aforesaid color extraction principle and that the color separation parameter and the data information are additionally stored in the respective document discriminating information of the document discriminating dictionary unit.

While an input unit 105' is provided for color constituent extraction, the input unit 105 may be used for that purpose.

Then, the operation of the document discriminating apparatus according to the embodiment will be described below for a registration step and a document discriminating step individually.
[Process in Registration Step]

Similarly to the control block diagram shown in Fig. 4, Fig. 9 shows the construction of a control block diagram when attention is paid in particular to the activation of the registration step in the document discriminating apparatus according to the present invention.

The control block shown in Fig. 9 in which attention is paid in particular to the activation of the registration step is constituted by an image storing memory 202, an image data cutting out unit 203, and a document discriminating dictionary unit 204. Then, the image storing memory 202 is designed to store data from the image reader 101 which is obtained when a document is read with the image reader 101. The control block of this embodiment is similar to the construction of the control block shown in Fig. 4 in that a hard disk 103 is connected thereto in which the data read from all documents is stored. However, the document discriminating apparatus according to the present invention is characterized further in that the color extracting unit and the color constituent separating unit are additionally provided.

The operation of the document discriminating apparatus in the registration step will be described below with reference to Fig. 10 showing a flowchart which explains operations at the time of activation of the registration step and Fig. 3.

Namely, as shown in Fig. 3, when image data of a document 106 (for example, an electric rate bill) is read at the image data reader 101 through the operation by the operator (step S1001), the image data so read is temporarily stored in the image data storing memory 202, and this image data is also stored in the hard disk 103, whereby the whole image data read at the image data reader 101 is stored (step S1002).

Note that the image data read at the image data reader 101 is indicated via the display 104, as shown in

Fig. 3 (step S1003).

Here, in a case where the image data stored in the image data storing memory 202 and the hard disk 103 is the first image data that has ever been read related to a document, the document discriminating information is stored in the document discriminating dictionary unit 204 as will be described below.

Namely, while referring to the display 104, the operator operates the input unit 105 for designation of areas of unique information to be extracted relative to the image cutting out unit 203 (step S1004). For example, as shown in Fig. 3, the 'ELECTRIC RATE', which is the first document discriminating information, is designated by surrounding it with a frame.

Then, since the position (X_0 , Y_0) and size of the document discriminating information can be specified when the document discriminating information is designated, the image cutting out unit 203 automatically cuts out image data for extraction which corresponds to the unique information described on the document from the image data storing memory 202 (step S1005).

Next, the cut out image data for extraction is sent to the color constituent extracting unit 213, where color information is analyzed from the cut out image data, and a density distribution related to each color constituent shown in Fig. 8 is prepared following the aforesaid color constituent extraction principle, and the distributed state is then displayed on the display 104 (step S1006). The operator can designate display conditions through the input unit 105'.

Then, the operator looks at the density distributions of the respective color constituents shown on the display 104 and selects a characteristic color constituent from the respective color constituents. The operator then designates the color constituent and a boundary a related to the color constituent relative to the color constituent extracting unit 213 by operating

the input unit 105'. The color constituent extracting unit 213 receives the designation and determines a color separation parameter (step S1007).

5 The determined color separation parameter is then sent to the color constituent separating unit 214. Here, the color constituent separating unit 214 extracts data information from the cut out image data following the color separation parameter (step S1008).

10 Next, the position information, size information, color separation parameter and the data information related to the discriminating information which are prepared as described above are registered in the document discriminating information column of the corresponding document in the document discriminating
15 dictionary unit 204 (step S1009).

In addition, when designating a plurality of extraction objects, the operations from step S1004 to step S1009 are repeated for each extraction object to obtain discriminating information. The discriminating
20 information so obtained for the plurality of extraction objects is registered for each document discriminating information of the corresponding document in the document discriminating dictionary unit 204.

Referring to the aforesaid Fig. 3, for example, the
25 electric rate indicating the details of the amount of money to be paid which is described on the bill is designated as first unique information by the operation of the input unit 105 by the operator, and 'FUJI ICHIRO' indicating the designation of a payer of the bill is
30 designated as second unique information. Then, at the image data cutting out unit 203, the position information, size information, color separation parameter and data information in the first unique information are stored, and the position information, size information,
35 color separation parameter and data information in the second unique information are extracted and stored in an area 204a of the document discriminating dictionary unit

204.

Thus, respective information related to a plurality of pieces of unique information is also extracted for a document B, and document discriminating information obtained is then stored in an area 204b of the document discriminating dictionary unit 204 as document discriminating information of the document type B having an ID number '0102' affixed thereto.

Note that in this document discriminating apparatus the image data cut out at the image data cutting out unit 203 is used only for discrimination of the document.

[Operation in Document Discriminating Step]

Similar to the control block diagram shown in Fig. 6, Fig. 11 shows a control block diagram when attention is paid to the activation of the document discriminating step.

The control block shown in Fig. 11 comprises an image storing memory 202, an image data cutting out unit 203, a document discriminating dictionary unit 204, a data comparison unit 205, a threshold setting unit 206 and a document determination unit 207, and the control block of Fig. 11 is similar to that of Fig. 6 in that the image storing memory 202 is designed to store data from the image reader 101 which is obtained when a document is read with the image reader 101. However, the control block shown in Fig. 11 is characterized in that a color constituent separating unit 214 is added to the document discriminating apparatus according to this embodiment.

While the color constituent extracting unit 213 is included in the construction of the control block shown in Fig. 9 which is used when the registration step is activated, the color constituent extracting unit 213 is not included in the construction of the control block shown in Fig. 11 since the discriminating information does not have to be registered in the latter control block.

The operation of the document discriminating

apparatus according to the embodiment of the present invention when the document discriminating step is activated will be described below using a flowchart for explaining the operation thereof when the document discriminating step is activated.

As has been described previously, in the verification step, whether or not the images of all the documents stored in the hard disk 103 can be specified using the document discriminating information registered in the document discriminating dictionary unit 204 is verified, and when the verification is completed, at the time of actual document discrimination, an operation which will be described below as a discriminating step for specifying the type of a document is performed.

Firstly, the image data of a certain document is read when the operator operates the image data reader 101. The image data so read is then temporarily stored in the image data storing memory 202 (step S1201).

Next, the image data cutting out unit 203 sequentially selects and reads out document discriminating information for each document sequentially in the order of the types of documents stored from the document discriminating dictionary unit 204 (step S1202).

Following this, the relevant data is cut out from the image data temporarily stored in the image data storing memory 202, based on the position information and size information which constitute the document discriminating information to be extracted in the first document type (step S1203).

Then, the color constituent separating unit 214 reads out a color separation parameter from the document discriminating information of the first document type in the document discriminating dictionary unit 204, and this color separation parameter is applied to the data cut out in step S1203. Here, the color constituent designated for this cut out image data is separated, and data information is extracted following the aforesaid

extraction principle (step S1204).

5 The data information prepared in step S1204 is sent to the data comparison unit 205, where the extracted data information is compared with the data information read out from the document discriminating information of the first document type, and the degree of matching with the data information is calculated (step S1205).

10 Here, the calculation of the degree of matching for one piece of document discriminating information in the first document type is completed. Next, the degree of matching is sequentially calculated for a plurality of pieces of document discriminating information, and such a calculation is performed for all the document discriminating information. Then, the degree of matching of all the discriminating information is compared with the degree of matching set at the threshold setting unit 206, and whether or not the degree of matching of all the discriminating information meets the judgment criteria is determined (step S1206).

20 In case the degree of matching of all the discriminating information meets the judgment criteria (Y), it means that the document read with the image reader 101 is specified as the first document type that has ever been read, and as this occurs, the specified document type is then recorded in a memory, not shown, within the control unit 201 as the result of the specification. On the contrary, the degree of matching of all the discriminating information does not meet the judgment criteria (N), since it means that no specification is implemented, no record is made in step S1207.

30 Thus, while the discriminating operation from the first document type stored in the document discriminating dictionary unit 204 is completed, such discriminating operations from all the document types stored in the document discriminating dictionary unit 204 are to be executed (step S1208).

Here, in case the discriminating operation has not yet been completed with respect to all the document types stored in the document discriminating dictionary unit 204 (N), the operations from steps S1203 to S1208 are repeated until the discriminating operation has been completed with respect to all the document types stored in the document discriminating dictionary unit 204.

In case the discriminating operation has been completed with respect to all the documents (Y), whether or not the document type recorded in step S1207 exists is determined (step S1209). In case there are recorded one or more document types (Y), if there is one, the document type thereof, and if there are recorded a plurality of document types, then the document type which is closest is informed the operator of or displayed on the display 104 (step S1210).

On the contrary, in case there is recorded no document type in step S1209, in other words, in case there is no specific document type in the step (N), that the document read with the image reader 101 cannot specify the document type is informed or displayed.

Thus, when the discrimination of the document read with the image data reader 101 is completed, the following document needed to be discriminated is read with the image data reader 101, and the discriminating operations are repeated.

Effectiveness of the Invention

As has been described heretofore, the document discriminating apparatus according to the present invention comprises the image data reader 101, the image data storing memory 202, the hard disk 103, the image data cutting out unit 203, the document discriminating dictionary unit 204, the data comparison unit 205 and the document determination unit 207, and furthermore, the color constituent extracting unit and the color constituent separating unit are additionally provided. Then, the color constituents of the data cut out as the

document discriminating information from the image data
read from the document are separated and the color
constituent indicating the characteristics is designated
from the density distributions of the respective color
5 constituents, so that the color separation parameter can
be set.

The color information can be replaced with
monochromatic information which can easily compare and
discriminate color information by the fact that the color
10 separation parameter can be set. Thus, thin spots in
color at the area to be extracted and variation in tone
between documents at the time of printing and reading can
be properly dealt with, and furthermore, the influence of
the background color can be eliminated. Owing to this,
15 the accuracy of the document discriminating information
can further be improved.

In discriminating documents, the type of a document
whose image data is read with the image data reader 101
can automatically be discriminated, and the accuracy of
20 the data cut out from the image data can be improved by
the color separation parameter based on the color
constituent extraction. Even if there exist a plurality
of types of documents to be read at the image reader 101,
the operator can process the documents without paying
25 attention to the definition of each document, whereby the
efficiency of the processing job can be improved. In
addition, there is no need to describe an ID number for
discriminating the document itself, whereby general types
of documents can be used, thus the document
30 discriminating apparatus of the present invention being
able to be applied to existing systems with ease.

Additionally, in registering the document
discriminating information into the document
discriminating dictionary unit 204, as the required
35 document discriminating information can automatically be
taken in when the operator so designates while looking at
the image data of the document to be registered which is

shown on the display 104, the dictionary for discriminating documents can easily be prepared, thereby making it possible to attempt to increase the efficiency of the document processing job.

5 Furthermore, as the document discriminating
information that is extracted through the designation of
the discriminating information by the operator can be
designated at the plurality of locations of the document,
the document can be specified with higher accuracy when
10 compared with the document discriminating information
which is designated at the single location.

What is claimed is:

1. An apparatus for discriminating a document with a data information, said apparatus comprising:

5 image reading means for reading image data from a document prepared in an optional format;

image data cutting out means for cutting out data corresponding to a designated specified portion of said document from said image data read by said image reading means;

10 color constituent extracting means for analyzing color constituents of said image data cut out by said cutting out means and setting a color separation parameter for a specific color constituent; and

15 color constituent separating means for producing said data information for said specific portion from said image data cut out based on said color separation parameter from said color constituent extracting means.

20 2. An apparatus for discriminating a document as set forth in Claim 1, wherein said color separation parameter set by said color constituent extracting means is stored in a document discriminating dictionary unit together with said data information.

25 3. An apparatus for discriminating a document as set forth in Claim 2, further comprising a document determination means for comparing for determination said data information prepared from image data obtained by reading a document to be discriminated based on said color separation parameter with said data information stored in said document discriminating dictionary unit.

30 4. A method for discriminating a document prepared in an optional format based on image data read from said document, said method comprising the steps of:

35 cutting out image data corresponding to a designated specified portion of said document;

analyzing color constituents of said image data so cut out, selecting a specific color constituent

and setting a color separation parameter; and
preparing data information for said
specified portion from said cut out image data based on
said color separation parameter; whereby

5 said document is discriminated by said
data information.

5. A method for discriminating a document as set
forth in Claim 4, wherein said color constituent is
analyzed with three primary colors of color, one of said
10 three primary colors is selected as said specific color
constituent, and said color separation parameter is
determined based on density distributions of said three
primary colors.

6. A method for discriminating a document as set
15 forth in Claim 4, wherein said color separation parameter
is stored in the document discriminating dictionary unit
together with said data information.

7. A method for discriminating a document as set
forth in Claim 6, wherein data information is prepared
20 from image data obtained by reading a document to be
discriminated based on said color separation parameter,
and wherein said data information so prepared is compared
for determination with said data information stored in
said document discriminating dictionary unit.

ABSTRACT

The invention relates to a document discriminating apparatus and a discriminating method for use in
5 processing documents at financial institutions. A characteristic portion inherent in an optional format is cut out from image data read from a document in the optional format. Color constituents of the cut out image data are analyzed, and a color constituent exhibiting
10 characteristics is selected from the constituents, and a color separation parameter is set for the selected color constituent. Data information is prepared which is related to the image data cut out based upon the color separation parameter. On the other hand, data
15 information is prepared from image data obtained by reading a document to be discriminated based on the color separation parameter. Then, the data information is compared for determination with the data information stored in the document discriminating dictionary unit,
20 whereby the document is discriminated.

Fig.1

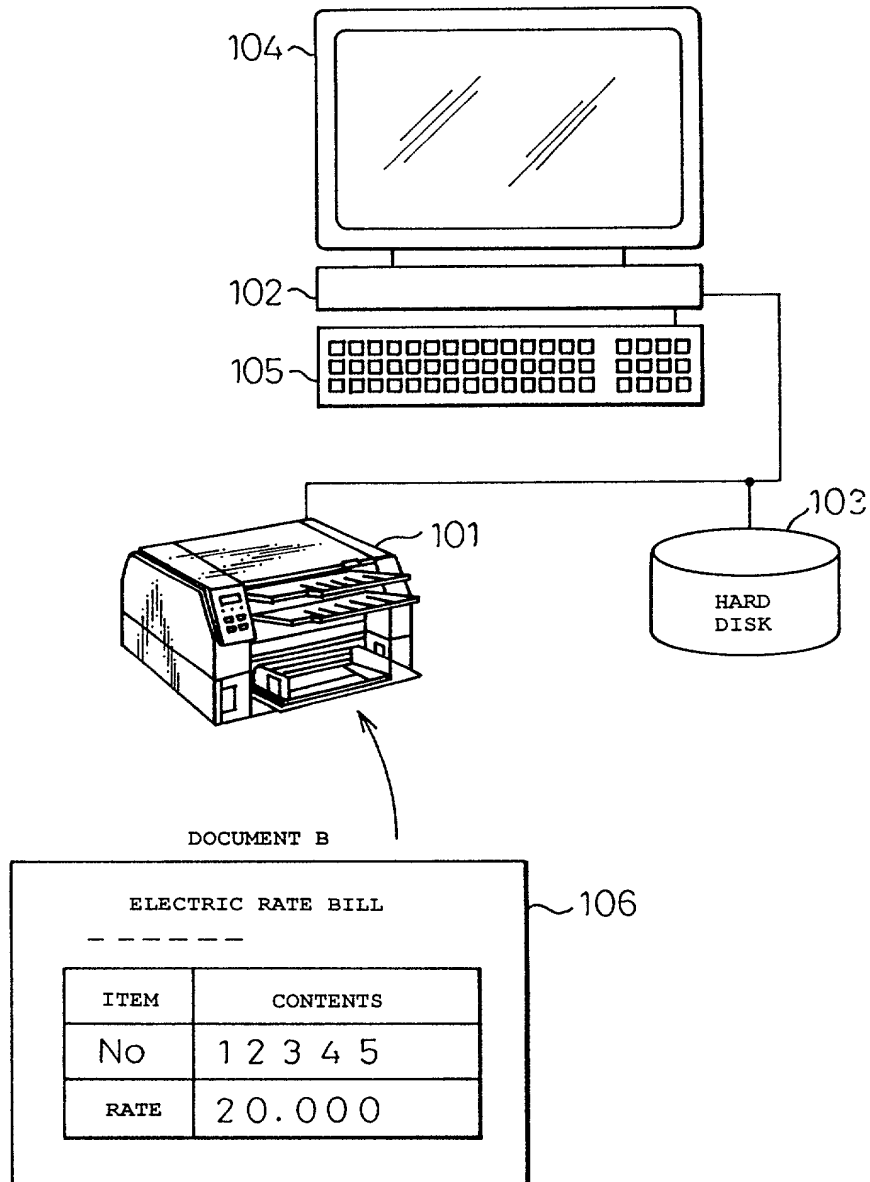


Fig.2

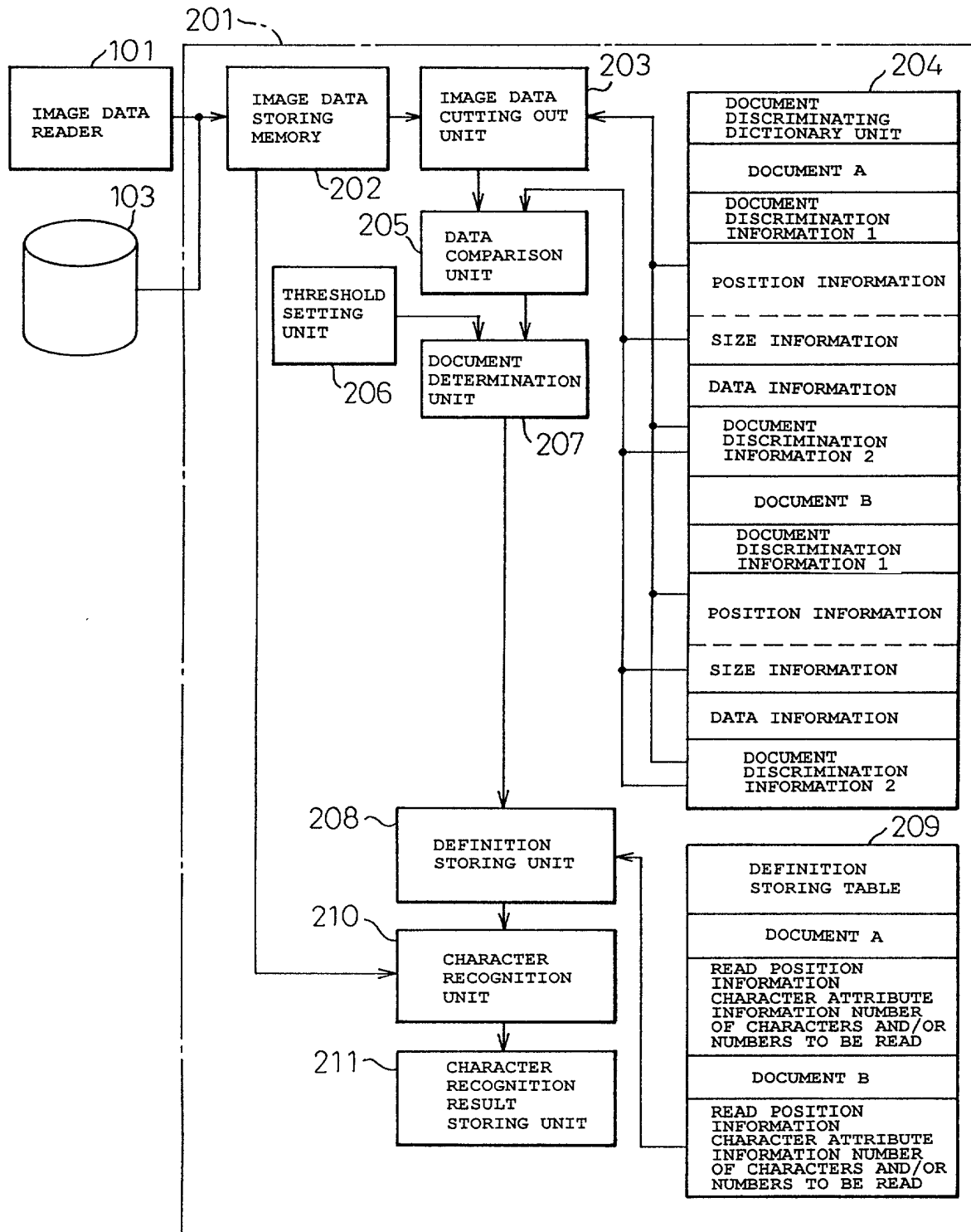


Fig.3

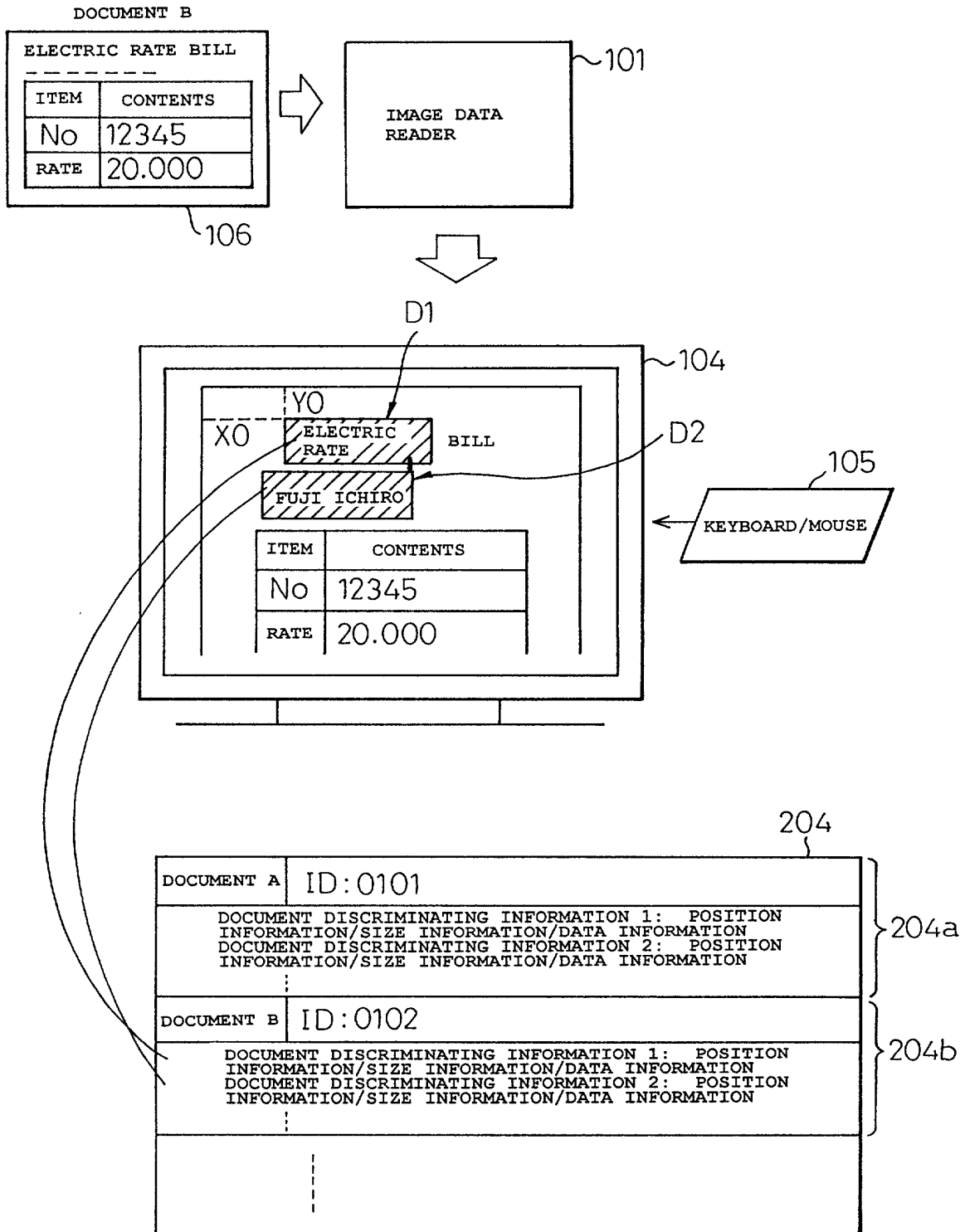


Fig.4

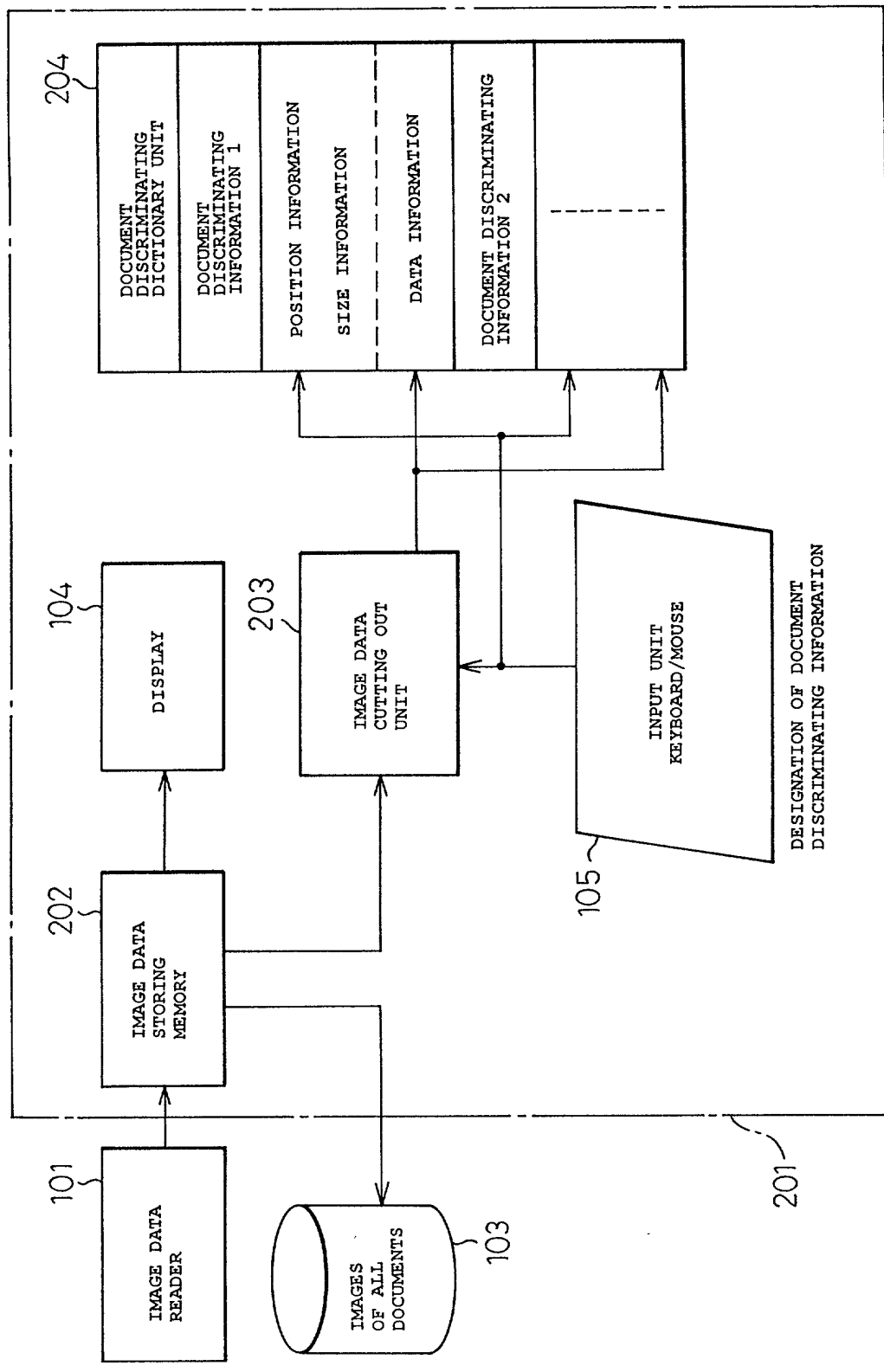


Fig.5

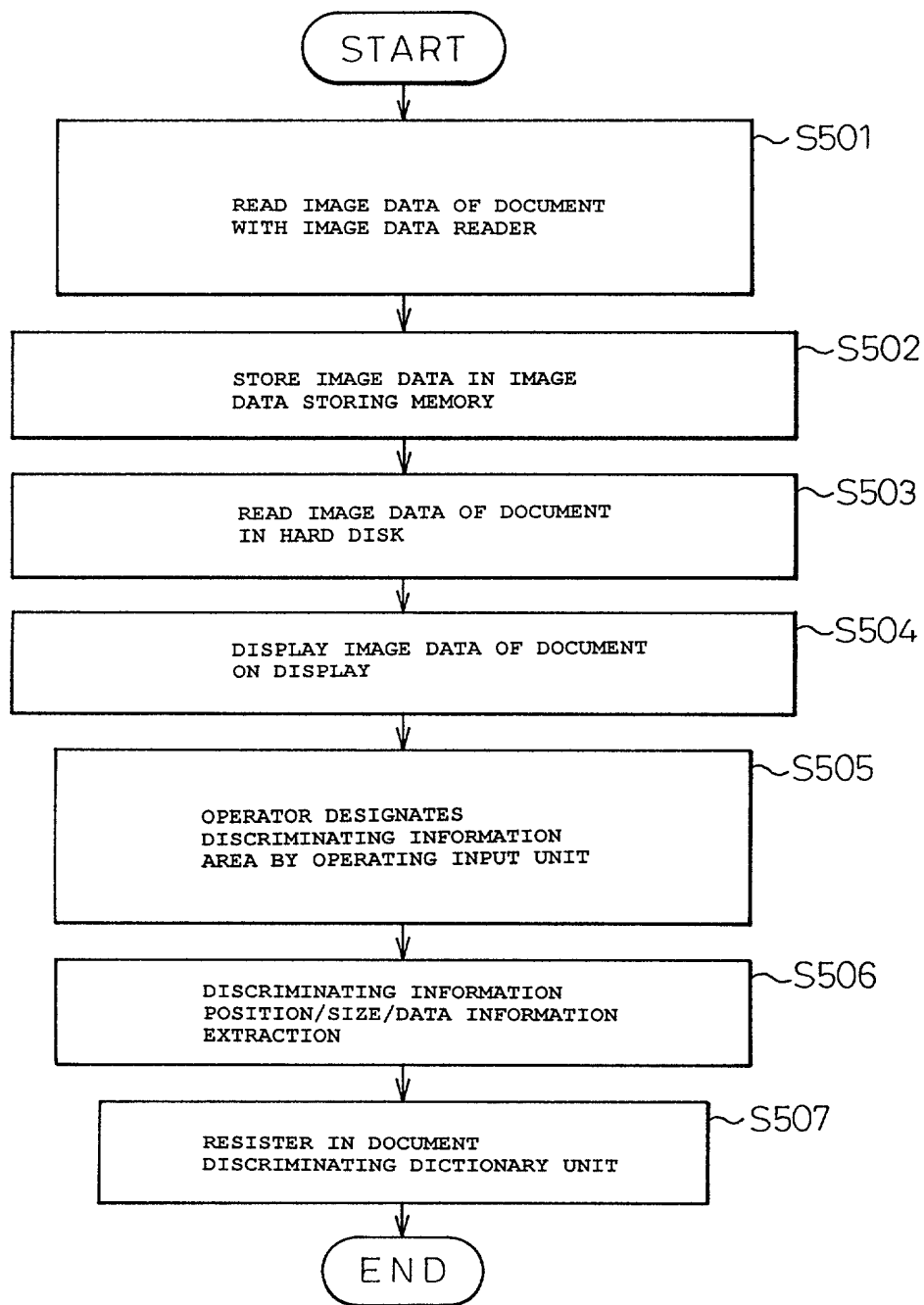


Fig.6

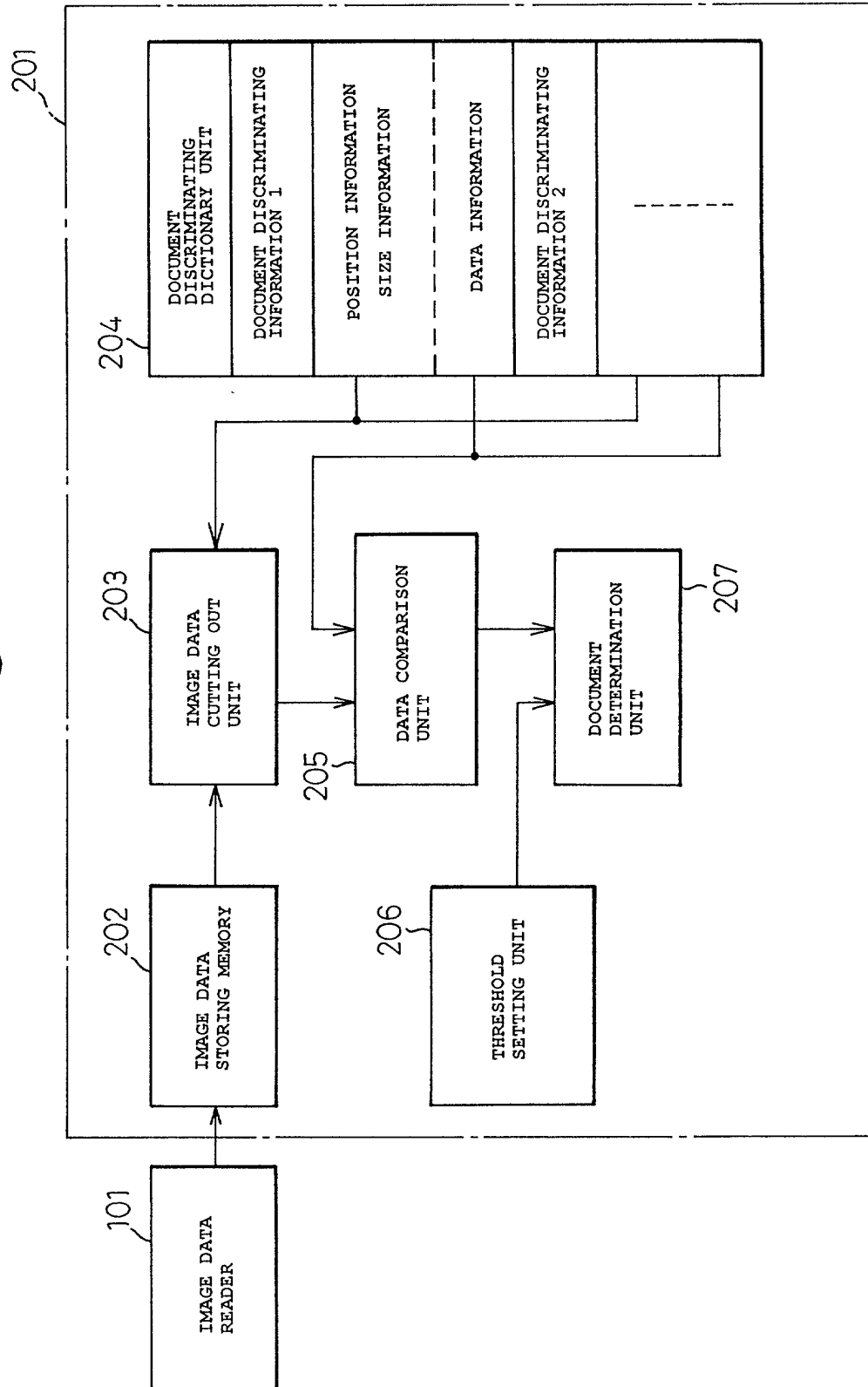


Fig.7

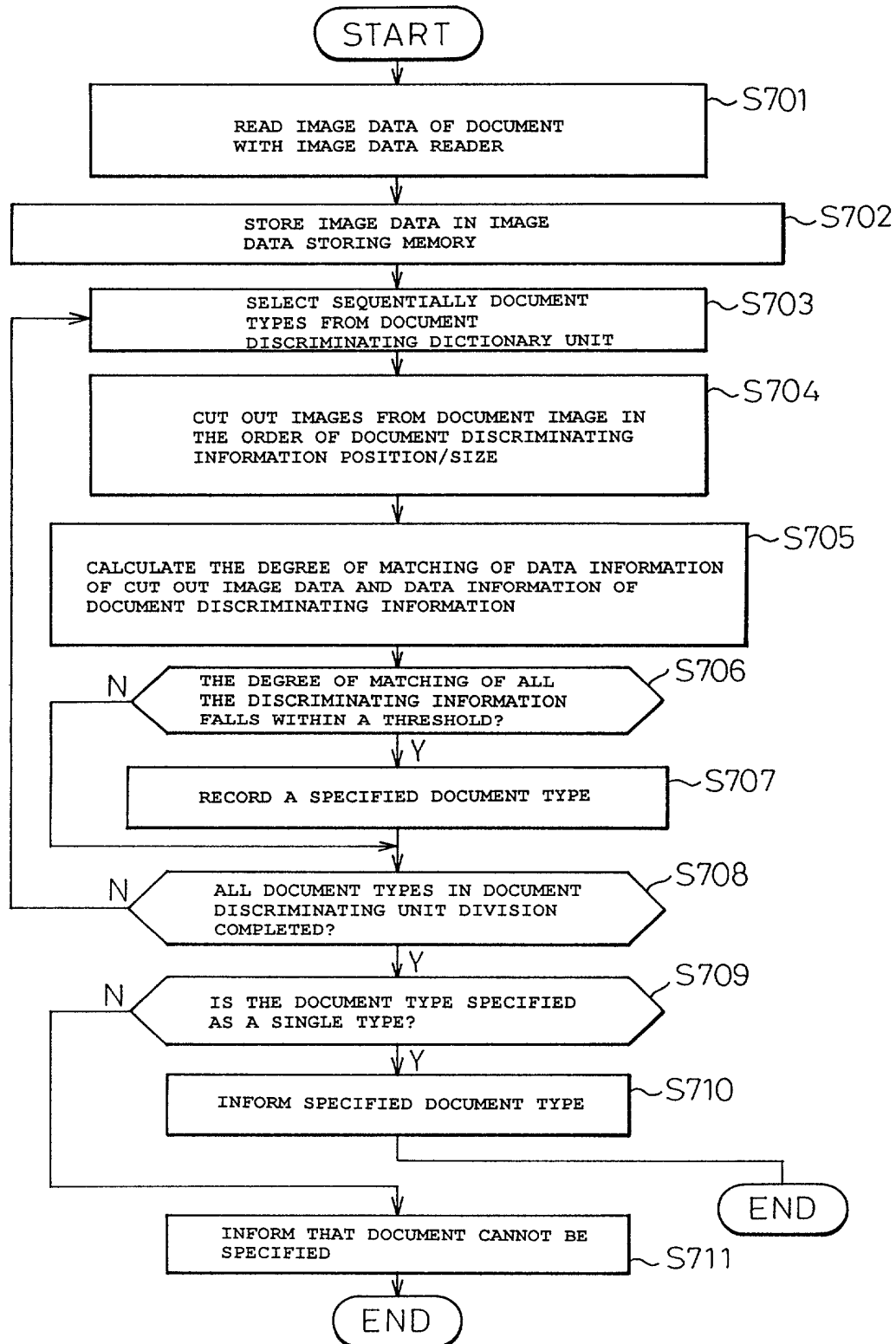


Fig.8

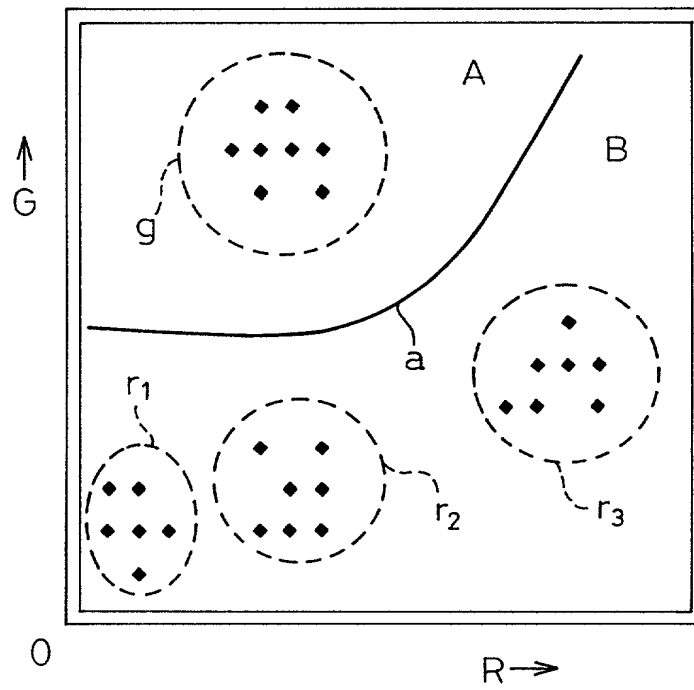


Fig.9

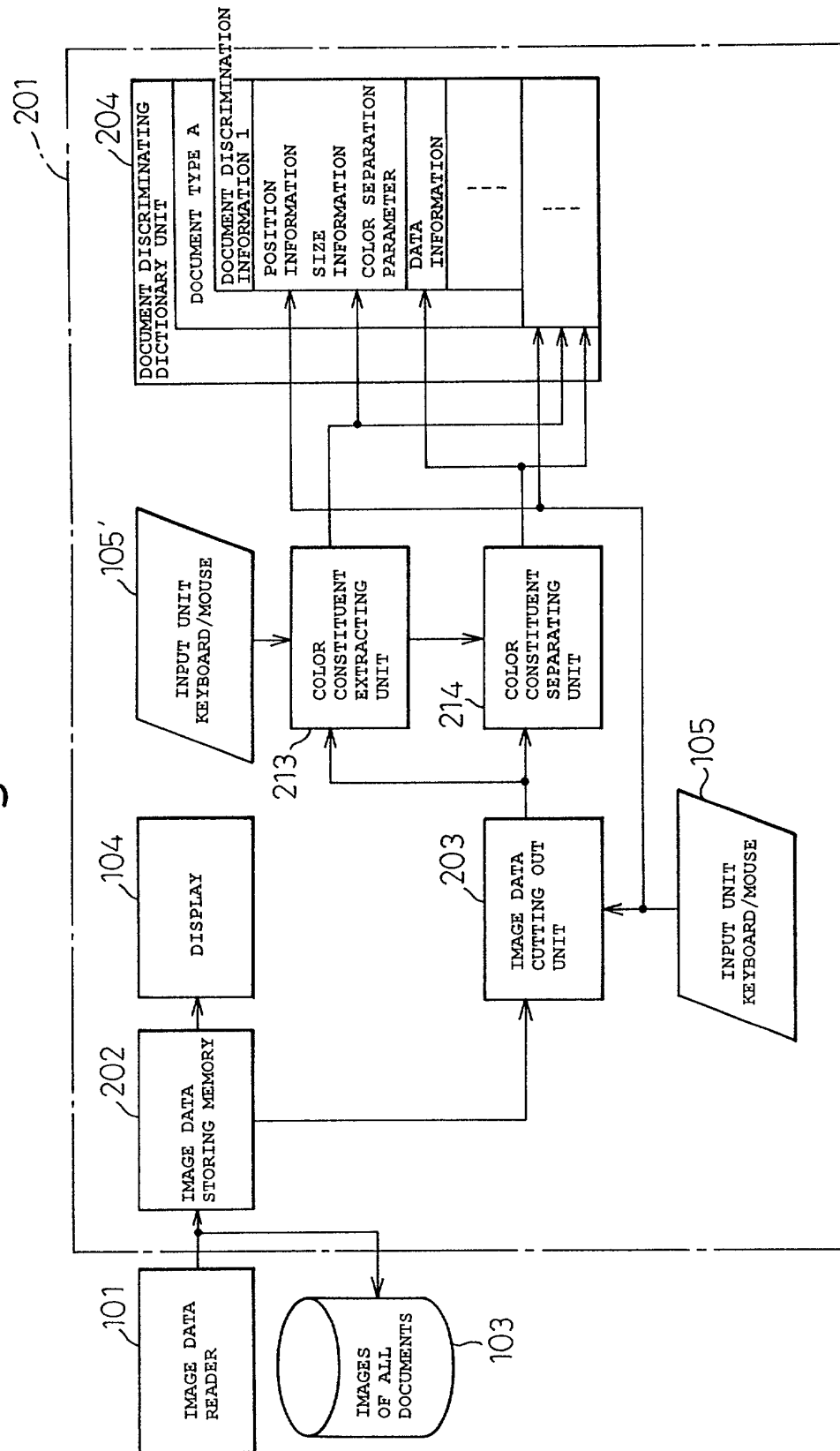


Fig.10

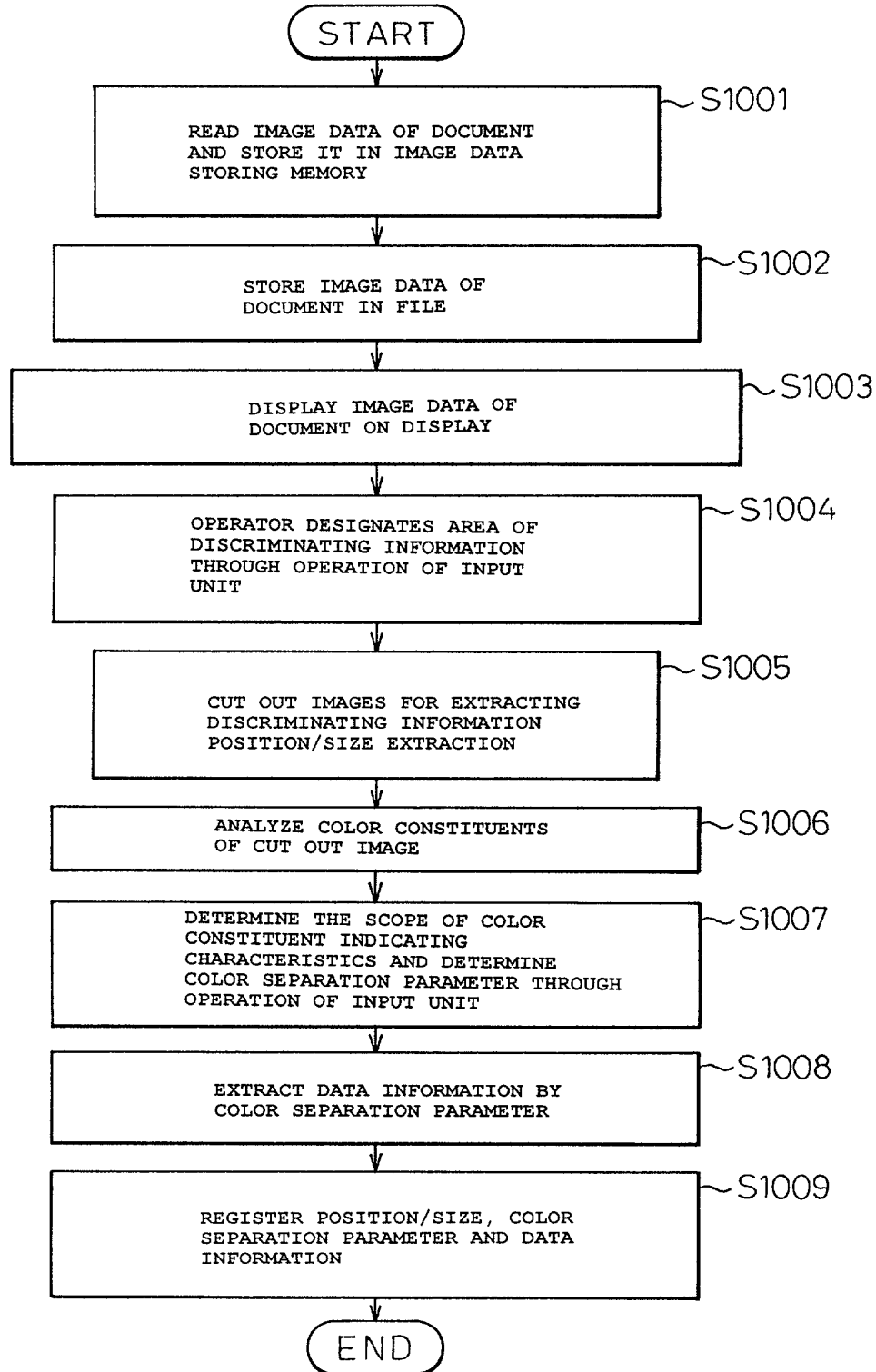


Fig.11

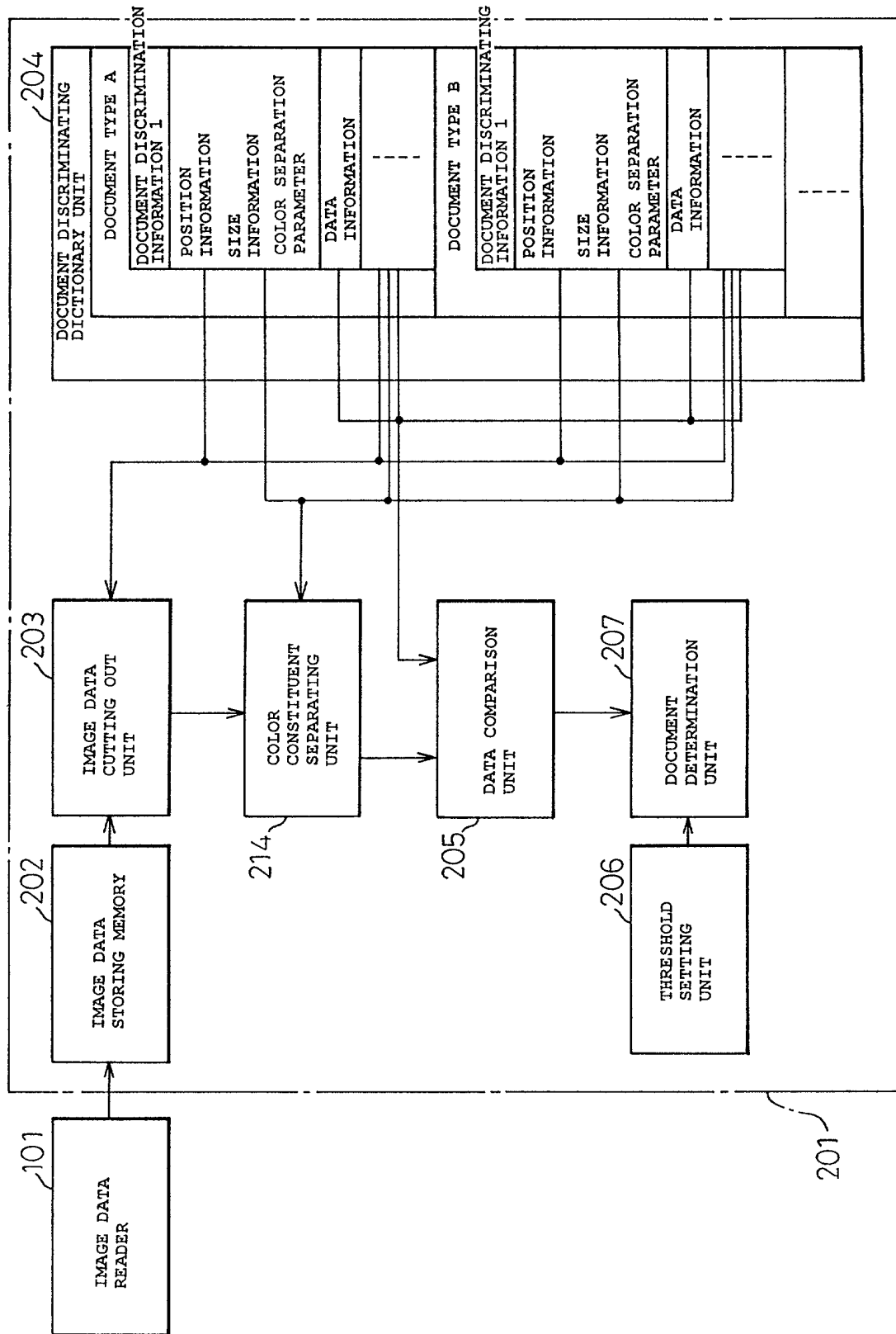
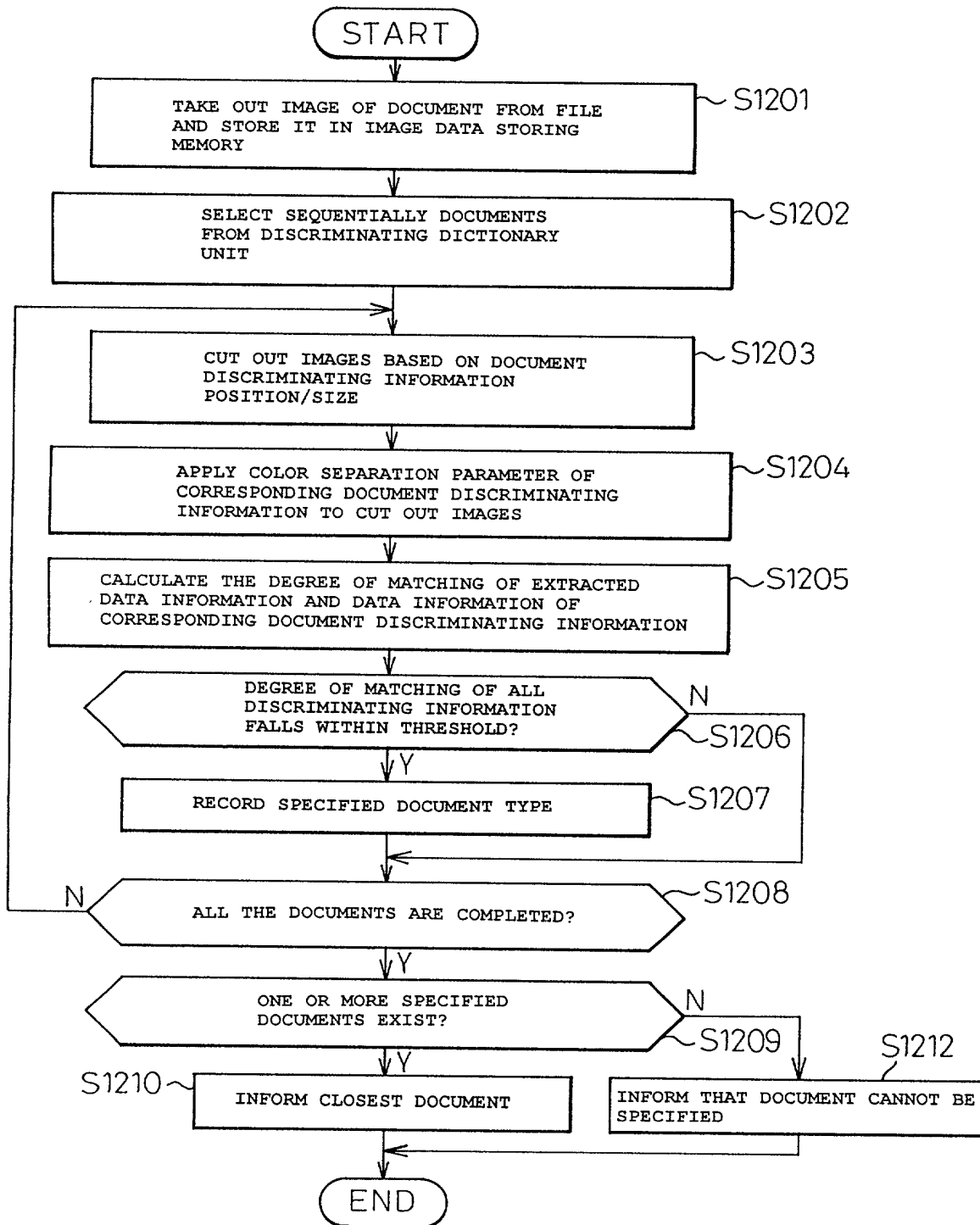


Fig.12



Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Declaration and Power of Attorney For Patent Application

特許出願宣言書及び委任状

Japanese Language Declaration

日本語宣言書

下記の氏名の発明者として、私は以下の通り宣言します。

As a below named inventor, I hereby declare that:

私の住所、私書箱、国籍は下記の私の氏名の後に記載された通りです。

My residence, post office address and citizenship are as stated next to my name.

下記の名称の発明に関して請求範囲に記載され、特許出願している発明内容について、私が最初かつ唯一の発明者（下記の氏名が一つの場合）もしくは最初かつ共同発明者であると（下記の名称が複数の場合）信じています。

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

APPARATUS AND METHOD FOR
DISCRIMINATING DOCUMENTS

上記発明の明細書（下記の欄でx印がついていない場合は、本表に添付）は、

the specification of which is attached hereto unless the following box is checked:

☐ 月 日に提出され、米国出願番号または特許協定条約
国際出願番号を _____ とし、
(該当する場合) _____ に訂正されました。

☐ was filed on _____
as United States Application Number or
PCT International Application Number
_____ and was amended on
_____ (if applicable).

私は、特許請求範囲を含む上記訂正後の明細書を検討し、内容を理解していることをここに表明します。

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

私は、連邦規則法典第37編第1条56項に定義されるとおり、特許資格の有無について重要な情報を開示する義務があることを認めます。

I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56.

Under the Paperwork Reduction Act of 1995 no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Japanese Language Declaration

(日本語宣言書)

私は、米国法典第35編119条(a)-(d)項又は365条(b)項に基づき下記の、米国以外の国の少なくとも一か国を指定している特許出願条約365(a)項に基づく国際出願、又は外国での特許出願もしくは発明者証の出願についての外国優先権をここに主張するとともに、優先権を主張している、本出願の前に出願された特許または発明者証の外国出願を以下に、枠内をマークすることで、示しています。

Prior Foreign Application(s)

外国での先行出願

(Number) (番号)	(Country) (国名)
(Number) (番号)	(Country) (国名)

I hereby claim foreign priority under Title 35, United States Code, Section 119 (a)-(d) or 365(b) of any foreign application(s) for patent or inventor's certificate, or 365(a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or PCT International application having a filing date before that of the application on which priority is claimed.

Priority Not Claimed

優先権主張なし

(Day/Month/Year Filed) (出願年月日)
(Day/Month/Year Filed) (出願年月日)

私は、第35編米国法典119条(e)項に基づいて下記の米国特許出願規定に記載された権利をここに主張いたします。

(Application No.) (出願番号)	(Filing Date) (出願日)
-----------------------------	------------------------

I hereby claim the benefit under Title 35, United States Code, Section 119(e) of any United States provisional application(s) listed below.

(Application No.) (出願番号)	(Filing Date) (出願日)
-----------------------------	------------------------

私は、下記の米国法典第35編120条に基づいて下記の米国特許出願に記載された権利、又は米国を指定している特許協力条約365条(c)に基づき権利をここに主張します。また、本出願の各請求範囲の内容が米国法典第35編112条第1項又は特許協力条約で規定された方法で先行する米国特許出願に開示されていない限り、その先行米国出願書提出日以降で本出願書の日本国内または特許協力条約国際提出日までの期間中に入手された、連邦規則法典第37編1条56項で定義された特許資格の有無に関する重要な情報について開示義務があることを認識しています。

I hereby claim the benefit under Title 35, United States Code, Section 120 of any United States application(s), or 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of Title 35, United States Code Section 112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of application.

International Application

No. PCT/JP99/05398

September 30, 1999

Pending

(Application No.) (出願番号)	(Filing Date) (出願日)	(Status: Patented, Pending, Abandoned) (現況: 特許許可済、係属中、放棄済)
-----------------------------	------------------------	---

(Application No.) (出願番号)	(Filing Date) (出願日)	(Status: Patented, Pending, Abandoned) (現況: 特許許可済、係属中、放棄済)
-----------------------------	------------------------	---

私は、私自身の知識に基づいて本宣言書中で私が行なう表明が真実であり、かつ私の入手した情報と私の信じることに基づき表明が全て真実であると信じていること、さらに故意になされた虚偽の表明及びそれと同等の行為は米国法典第18編第1001条に基づき、罰金または拘禁、もしくはその両方により処罰されること、そしてそのような故意による虚偽の声明を行えば、出願した、又は既に許可された特許の有効性が失われることを認識し、よってここに上記のごとく宣言を致します。

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

Japanese Language Declaration

(日本語宣言書)

委任状： 私は下記の発明者として、本出願に関する一切の手続きを特許庁事務局に対して遂行する弁護士または代理人として、下記の者を指名いたします。（弁護士、または代理人の氏名及び登録番号を明記のこと）

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith (list name and registration number)

手数料

And I hereby appoint as principal attorneys: David T. Nikaido, Reg. No. 22,663; Charles M. Marmelstein, Reg. No. 25,895; George E. Oram, Jr., Reg. No. 27,931; Robert B. Murray, Reg. No. 22,980; E. Marcie Emas, Reg. No. 32,131; Douglas H. Goldhush, Reg. No. 33,125; Monica Chin Kitts, Reg. No. 36,105; Richard J. Berman, Reg. No. 39,107; King L. Wong, Reg. No. 37,500; Karen K. Costantino, Reg. No. 35,107; James A. Poulos, III, Reg. No. 31,714; Patrick D. Muir, Reg. No. 37,403; Sharon N. Klesner, Reg. No. 36,335; and Murat Ozgu, Reg. No. 44,275; Bradley D. Goldizen, Reg. No. 43,637; and N. Alexander Nolte, Reg. No. 45,689.

直接電話連絡先：（名前及び電話番号）

Please direct all communications to the following address:
ARENT FOX KINTNER PLOTKIN & KAHN, PLLC
1050 Connecticut Avenue, N.W., Suite 600
Washington, D.C. 20036-5339
Tel: (202) 857-6000; Fax: (202) 857-6395

第一または第一発明者名

Full name of sole or first inventor

発明者の署名

日付

Inventor's signature

Date

住所

Residence

国籍

Citizenship

私書箱

Post Office Address

c/o FUJITSU LIMITED, 1-1, Kamikodanaka
4-chome, Nakahara-ku, Kawasaki-shi,
Kanagawa 211-8588 Japan

第二共同発明者名

Full name of second joint inventor, if any

第二共同発明者の署名

日付

Second inventor's signature

Date

住所

Residence

国籍

Citizenship

私書箱

Post Office Address

c/o FUJITSU LIMITED, 1-1, Kamikodanaka
4-chome, Nakahara-ku, Kawasaki-shi,
Kanagawa 211-8588 Japan

(第三以降の共同発明者についても同様に記載し、署名をする)

(Supply similar information and signature for third and subsequent joint inventors.)

第三共同発明者名		300 Full name of third joint inventor, if any	
第三共同発明者の署名		Kazunori Yamamoto	
日付	Third inventor's signature	Date	February 8, 2002
住所	Residence	Kawasaki, Japan JPK	
国籍	Citizenship	Japanese	
私書箱	Post Office Address	c/o FUJITSU LIMITED, 1-1, Kamikodanaka 4-chome, Nakahara-ku, Kawasaki-shi, Kanagawa 211-8588 Japan	
第四共同発明者名		400 Full name of fourth joint inventor, if any	
第四共同発明者の署名		Shinichi Eguchi	
日付	Fourth inventor's signature	Date	February 8, 2002
住所	Residence	Kawasaki, Japan JPK	
国籍	Citizenship	Japanese	
私書箱	Post Office Address	c/o FUJITSU LIMITED, 1-1, Kamikodanaka 4-chome, Nakahara-ku, Kawasaki-shi, Kanagawa 211-8588 Japan	
第五共同発明者名		Full name of fifth joint inventor, if any	
第五共同発明者の署名		Fifth inventor's signature	
日付		Date	
住所	Residence		
国籍	Citizenship		
私書箱	Post Office Address		
第六共同発明者名		Full name of sixth joint inventor, if any	
第六共同発明者の署名		Sixth inventor's signature	
日付		Date	
住所	Residence		
国籍	Citizenship		
私書箱	Post Office Address		